DESDE 2013
https://repository.uaeh.edu.mx/revistas/index.php/icbi/issue/archive
Pädi Boletín Científico de Ciencias Básicas e Ingenierías del ICBI
ISSN: 2007-6363

Publicación Semestral Pädi Vol. 11 No. 21 (2023) 1–7

Modelos ARIMA para el análisis sistematizado de criptomonedas
ARIMA models for the systematic analysis of cryptocurrencies

G. Vital-Godinez[a,b,*], J. Quiterio-Romero[b], P. Miranda-Romagnoli[b], C. A. Soto-Campos[b]

[a]*Área Académica de Economía, Universidad Autónoma del Estado de Hidalgo, 42184, Pachuca, Hidalgo, México.*
[b]*Área Académica de Matemáticas y Física, Universidad Autónoma del Estado de Hidalgo, 42184, Pachuca, Hidalgo, México.*

Resumen

Se realiza un análisis ARIMA de las series de tiempo correspondientes al rendimiento de Bitcoin. Para ello se utiliza el software R. Se hace un estudio comparativo de diferentes modelos ARIMA para modelar el comportamiento del rendimiento de Bitcoin en el periodo comprendido del día 1 de enero de 2020 hasta el 31 de diciembre de 2020. Finalmente, se efectúa un análisis de bondad de ajuste para comprobar cuál de los modelos reproduce mejor los datos reportados.

Palabras Clave: ARIMA, Bitcoin, Pronóstico, Series de Tiempo.

Abstract

An ARIMA analysis of the time series corresponding to the performance of Bitcoin is performed. For this, the R software is used. A comparative study of different ARIMA models is made to model the performance behavior of Bitcoin in the period from January 1, 2020 to December 31, 2020. Finally, a goodness-of-fit analysis is carried out to verify which of the models best reproduces the reported data.

Keywords: ARIMA, Bitcoin, Forecast, Time Series.

1. Introducción

Los modelos de promedio móvil integrado autorregresivo ARIMA, del acrónimo del inglés Autoregressive Integrated Moving Average, constituyen una técnica de análisis de series de tiempo que permiten conocer la dependencia de una cierta variable en función de otras. El objetivo de este modelo es pronosticar valores futuros de la serie en relación con cierto número de variables conocidas. Existen varios trabajos modernos que utilizan este enfoque para analizar el comportamiento de las series de tiempo asociadas a criptomonedas. En (Alahmari, 2019) se ha investigado la predicción del precio de bitcoin utilizando un modelo ARIMA, se han procesado los datos de manera que sean estacionarios y después se ha buscado el modelo ARIMA que minimiza el error cuadrático medio de la predicción (MSE). Los resultados que se presentan indican que la predicción del precio del bitcoin da grandes valores de MSE debido a la vulnerabilidad de su precio, confirmando de esta manera que el modelo ARIMA aún se puede usar para la predicción de precios en subperíodos del intervalo de tiempo en donde los datos tienen una tendencia única. En (Roy *et al.*, 2018) también se presenta un modelo para predecir el precio del bitcoin en el mercado aplicando datos de los años 2013-2017 basados en un modelo ARIMA. Finalmente podemos decir que el enfoque de los modelos ARIMA permanece vigente debido a que se basa en una metodología sencilla que permite hacer predicciones con una gran economía de cálculos.

2. Marco Teórico

El análisis principal del presente trabajo de investigación se sustenta en los modelos estadísticos ARIMA, el cual es un modelo estadístico que se basa en el uso de series de tiempo para encontrar patrones con el fin de realizar predicciones a futuro. Uno de los puntos importantes respecto al análisis es el uso de series de tiempo, por lo que se profundizará más en ambos temas con el fin de dar una vista general de la motivación de su uso y cómo comparar este interés con los resultados obtenidos.

*Autor para correspondencia: vi397509@uaeh.edu.mx
Correo electrónico: vi397509@uaeh.edu.mx (G. Vital-Godinez), qu419144@uaeh.edu.mx (J. Quiterio-Romero), pmiranda@uaeh.edu.mx (P. Miranda-Romagnoli), csoto@uaeh.edu.mx (C. A. Soto-Campos).

Historial del manuscrito: recibido el 31/10/2022, última versión-revisada recibida el 15/01/2023, aceptado el 19/01/2023, en línea (postprint) desde el 01/02/2023, publicado el 05/07/2023. **DOI:** https://doi.org/10.29057/icbi.v11i21.10158

2.1. *Series de tiempo*

Para entender la idea que cimienta el análisis de series de tiempo, es más fácil partir de un ejemplo. Suponiendo que se tiene un banco de observaciones sobre alguna variable de interés con una marca temporal, se puede situar a las observaciones desde las temporalmente más antiguas a las más recientes, y con esta única relación realizar predicciones de cómo se comportará esta variable en el futuro. La idea es que las observaciones que se tienen del pasado mostrarán el comportamiento en el futuro de la variable. Esto remite a que las observaciones tendrán una relación entre ellas, pero no necesariamente debe ser así; lo que sí mostrará es si existe una tendencia y qué tan confiable es para realizar una predicción, para ilustrarlo se plantearán dos ejemplos.

El primer ejemplo consiste en lanzar una moneda repetidas veces para observar la tendencia de los datos obtenidos. En este caso se tienen dos posibles resultados: cara, con el valor de 1, y cruz, con el valor de -1; se puede suponer que cada tirada es un evento aislado en un tiempo y que la tirada n es más reciente que la tirada n-1 por lo que tendrían una relación temporal. De esta forma se buscará ver una tendencia en la Figura 1, esta figura son 100 lanzamientos simulados de una moneda, tomando un generador de números pseudo-aleatorios, en donde, si el número generado es par se le asigna el número 1, al que se toma como cara, y si el número generado es impar se le asigna el valor -1, al que se le considera como cruz. Como se puede observar en la Figura 1, no es posible ver una tendencia de los datos o una relación de los eventos en la tirada n con la n+m, donde m es cualquier valor mayor o igual que 1. En teoría para un muestra grande de tiradas se tendrá un 50 % de probabilidad de obtener cara, e igualmente, la misma probabilidad para obtener cruz, por lo que este es un buen ejemplo de un lugar donde no es aconsejable usar series de tiempo, para predecir futuras tiradas (dado que no hay una relación de los eventos del pasado con los del futuro, aunque claro, es poco probable que se repita la misma cara durante muchas tiradas consecutivas, pero, aun así, la probabilidad de cada evento individual es la misma.

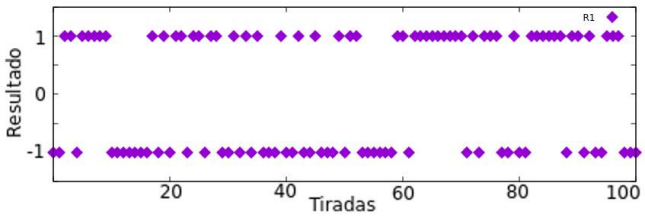


Figura 1: Tiros aleatorios de una moneda generados con números pseudo-aleatorios.

El segundo ejemplo considera los datos de la población económicamente activa en México desde el primer trimestre del 2005 hasta el segundo trimestre del 2022 (INEGI, 2022), estos se pueden ver representados en la Figura 2. Los datos están seccionados en trimestres, a los que se les refiere como primer periodo, segundo periodo, tercer periodo y cuarto periodo, respectivamente. Más allá de la conformación de los datos, es claro que los valores entre ellos varían relativamente poco, y aunque se aprecian caídas en los datos, son pequeñas y en su mayoría siguen una tendencia de crecimiento, exceptuando los valores del primer al segundo periodo del 2020, donde la caída es mayormente atribuida a la pandemia causada por el virus SARS-CoV-2. Para este fin ilustrativo, se tomarán solo los datos anteriores a la caída del segundo periodo del 2020; estos datos, como ya se mencionó, tienen una variación relativamente baja entre ellos, además de ello es posible ver una tendencia, donde se puede intentar relacionar los datos anteriores con los futuros; aunque no se realizará una prueba rigurosa en este momento, se puede apreciar que los datos separados por solo un periodo no varían mucho entre ellos, esto podría indicar que el siguiente periodo tampoco variará mucho con el anterior. Realizado esto de forma sucesiva, se podría construir una predicción a futuro. Este es un buen ejemplo en el cual se pueden usar las series de tiempo para intentar predecir los eventos futuros, lo cual daría una idea relativa de en qué lugares sería una buena opción aplicar métodos basados en series de tiempo para predecir eventos futuros, (Levendis, 2019).

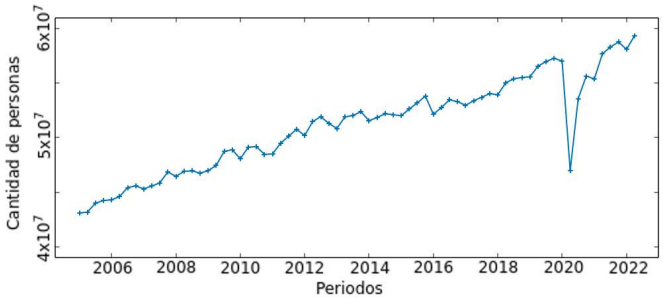


Figura 2: Datos de la población económicamente activa en México del primer trimestre del 2005 al segundo trimestre del 2022.

2.2. *Modelos ARIMA*

Dentro de los modelos basados en series de tiempo, se tienen los modelos ARIMA, que son una conformación de dos componentes que están inscritos en el nombre del modelo. AR se refiere a autorregresivo que modela relacionando los valores más viejos temporalmente para predecir los valores futuros y la parte MA que se refiere a medias móviles, la cual se basa en modelar por medio de los errores del pasado.

Para poder aplicar un modelo ARIMA, es necesario que se cumpla el supuesto de que la serie es estacionaria, significando que posee una media constante; en otras palabras, que al aplicarle una diferenciación, los valores no disten en gran medida del origen o que estos, a su vez, muestren un comportamiento simétrico alrededor del origen. Esta es una de las formas para saber si una serie es estacionaria; de igual forma es posible ver que tan correlacionados están los valores con el pasado. En caso de encontrarse con una serie no estacionaria, esta podría presentar problemas de raíces unitarias, resultando en que la función de autocorrelación disminuya muy lentamente; para ello existe una solución que no es automática, pero puede ser útil: diferenciar la serie hasta que esta sea estacionaria. Una vez hecho esto, es posible realizar un análisis por medio de modelos ARIMA (Levendis, 2019).

3. Metodología

En esta sección se discutirá la metodología empleada para la realización de los modelos econométricos, los cuales serán fundamentalmente del tipo ARIMA.

En primera instancia, se empezó recopilando y analizando la serie de BITCOIN-USD (dólar estadounidense) con datos por minuto para el año 2020 en su precio de cierre (último dato del día) desde el día 1 de enero de 2020 a las 00:01:00 hasta el 31 de diciembre de 2020 a las 23:59:00. Los datos fueron obtenidos de la página web (Bitstamp, 2020), la cual se encuentra en las referencias.

Con los modelos ARIMA, se deben seguir algunos pasos fundamentales ((Bakar y Rosbi, 2017)):

- Identificar el tipo de modelo ARIMA (p, d, q).
- Estimar el modelo respectivo.
- Realizar diagnósticos sobre el modelo estimado.
- Predicción.

Primero hay que enfocarse en las dos partes más importantes del modelo ARIMA. Desglosando los aspectos matemáticos, un modelo autorregresivo AR(p) tiene la siguiente forma:

$$\begin{aligned} X_t &= \beta_0 + \beta_1 X_(t-1) + \beta_2 X_(t-2) \\ &\quad + \ldots + \beta_p X_(t-p) + \epsilon_t, \end{aligned} \tag{1}$$

$$X_t = \beta_0 + \sum_{i=1}^{p} \beta_i X_{t-i} + \epsilon_t, \tag{2}$$

donde β_0 es una constante cualquiera (que funge como intercepto), β_i son los parámetros respectivos de cada dato X_{t-i} (los parámetros beta pueden tomar cualquier valor real $\Re$) y ϵ_t es el error estocástico del modelo (ruido blanco).

Asimismo, se tiene la sección de media móvil del modelo, MA(q), la cual está dada de la siguiente manera:

$$X_t = \mu + \epsilon_t + \theta_1 \epsilon_{t-1} + \theta_2 \epsilon_{t-2} + \ldots + \theta_q \epsilon_{t-q} \tag{3}$$

$$X_t = \mu + \epsilon_t + \sum_{i=1}^{q} \theta_i \epsilon_{t-i}, \tag{4}$$

siendo μ la media de la serie, θ_i son los parámetros del modelo y ϵ_k son los errores estocásticos del modelo, donde $k \in [t-q, t]$. Considerando las dos partes (la autorregresiva y la de media móvil), esto conduce a un modelo llamado ARMA, el cual tiene la siguiente forma:

$$X_t = \beta_0 + \sum_{i=1}^{p} \beta_i X_{t-i} + \epsilon_t + \mu + \sum_{i=0}^{q} \theta_i \epsilon_{t-i} + \epsilon_t. \tag{5}$$

No obstante, este aún no es el modelo deseado. Para ello, se procede a realizar una derivación a la ecuación (5) y se introduce un operador llamado Lag (L), el cual, en series de tiempo, tiene la propiedad de que $L^h X_t = X_{t-h}$, donde h siempre deberá ser un número entero. Ahora, sea α_i el parámetro de la parte autorregresiva después de una derivación, por lo que el modelo ARMA (p, q) pasará a ser ARMA (p', q). Con esta diferenciación, la ecuación (5) puede ser reescrita de la siguiente manera:

$$X_t - \alpha_1 X_{t-1} - \ldots - \alpha_{p'} X_{t-p'} = \epsilon_t + \theta_1 \epsilon_{t-1} + \ldots + \theta_q \epsilon_{t-q} \tag{6}$$

$$\left(1 - \sum_{i=1}^{p'} \alpha_i L^i\right) X_t = \left(1 + \sum_{i=1}^{q} \theta_i L^i\right) \epsilon_t. \tag{7}$$

En la ecuación (7), en el término $\left(1 - \sum_{i=1}^{p'} \alpha_i L^i\right)$ se asumirá que existe una raíz unitaria de multiplicidad d; una raíz unitaria significa que existe una parte de la serie que tiene una tendencia estocástica y, por ende, resulta impredecible; esta raíz unitaria será de la forma $(1 - L)$, por lo que la ecuación puede ser reescrita sacando la raíz unitaria:

$$\left(1 - \sum_{i=1}^{p'-d} \alpha_i L^i\right) (1 - L)^d X_t = \left(1 + \sum_{i=1}^{q} \theta_i L^i\right) \epsilon_t, \tag{8}$$

y definiendo $p = p' - d$, ϕ_i serán los nuevos parámetros autorregresivos y nuestro modelo estará dado por:



Figura 3: Serie del Bitcoin para el año 2020.

$$\left(1 - \sum_{i=1}^{p} \phi_i L^i\right) (1 - L)^d X_t = \left(1 + \sum_{i=1}^{q} \theta_i L^i\right) \epsilon_t. \tag{9}$$

4. Resultados

Empezando con los resultados del modelo ARIMA, se graficaron los datos del precio del BITCOIN (BTC) en dólares estadounidenses (USD) con respecto al tiempo transcurrido (minutos), del 1 de enero de 2020 al 31 de diciembre de 2020; como se puede apreciar en la Figura 3, la cual muestra una alta volatilidad a lo largo del año, por lo que la ejecución de ciertas pruebas es necesario.

Por ello, se hicieron dos pruebas: la función de autocorrelación (ACF por sus siglas en inglés -autocorrelation function-) y la función de autocorrelación parcial (PACF -partial autocorrelation function-), como se muestra en las figuras 4 y 5.

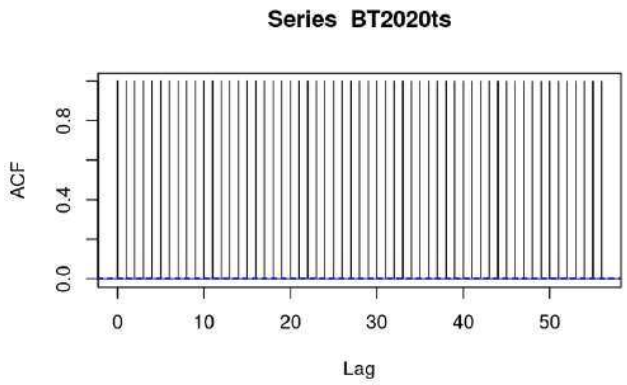


Figura 4: Función de autocorrelación, ACF.

Como se puede observar en la Figura 4 de la ACF, la serie está muy correlacionada con su pasado. Esto se debe a que la serie es claramente no estacionaria, por lo que se buscará que lo sea. Es por ello que se tomó la serie y se diferenció una vez, dejando la serie de la siguiente manera que se muestra en la Figura 6.

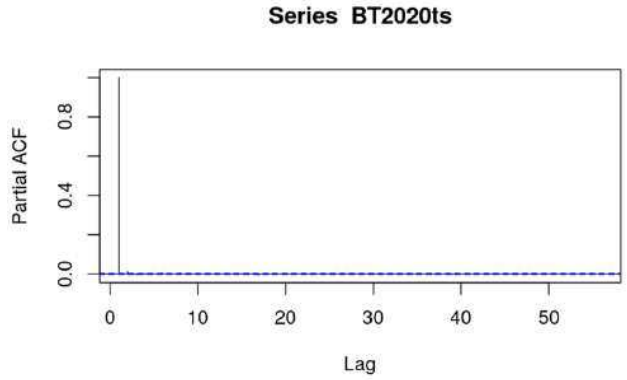


Figura 5: Función de autocorrelación parcial, PACF.

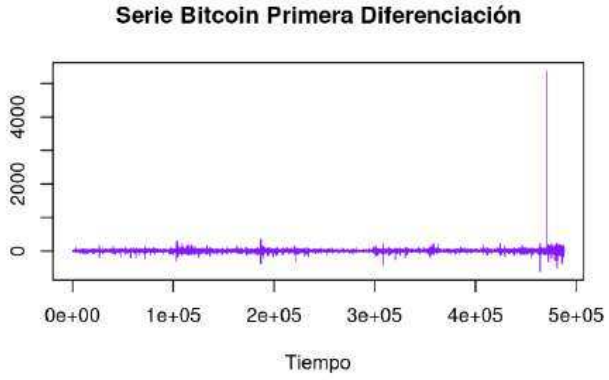


Figura 6: Serie Diferenciada.

Como se puede apreciar, contrastando con la serie original (mostrada en la Figura 3), ya los datos rondan un punto fijo, que en este caso es una media cercana al cero (0.0448 para ser exactos). Con esto, se obtuvo la ACF y PACF de la nueva serie diferenciada mostradas en las Figuras 7 y 8.

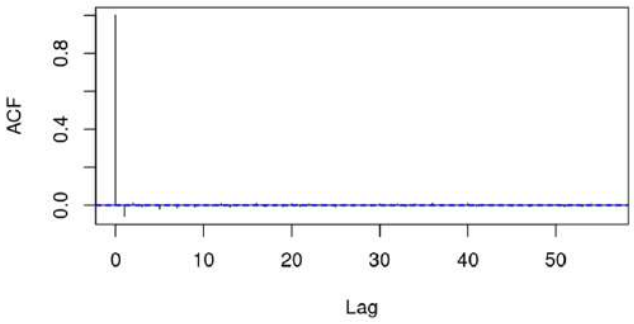


Figura 7: ACF con una diferenciación.

Con esto, se pudo determinar el modelo ARIMA: una diferenciación, dos términos autorregresivos y sin media móvil; o lo que es lo mismo, un modelo ARIMA (2,1,0). La parte autorregresiva se determinó del comportamiento de la ACF, teniendo un decaimiento geométrico, así como en la gráfica de la serie diferenciada se podía ver que los datos fluctuaban desde una media específica y el orden fue dado por el PACF donde se tomaron los dos primeros lags como significativos; la parte de media móvil se descartó por la gráfica de los datos, así como el decaimiento de la ACF; y la parte de diferenciación, simplemente fue dada para que los datos pudieran ser trabajados de manera más eficiente.

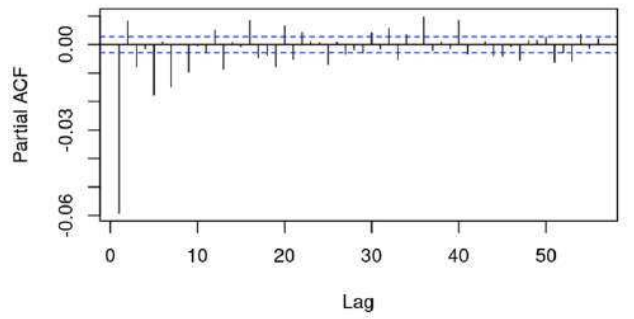


Figura 8: PACF con una diferenciación.

No obstante, se hizo una corrida de diversos modelos ARIMA, comparándolos a través del criterio de información Akaike, para ver cuál era efectivamente el mejor (Levendis, 2019). La corrida se hizo con todos los datos del año 2020, exceptuando los últimos veinte, buscando, posteriormente, comparar la predicción con los valores reales. En la Tabla1 se presentan los 67 modelos diferentes.

Una vez comparados los múltiples modelos, se obtuvo que el que tiene un AIC más bajo es el ARIMA (61,2,0), como se ve en la Tabla 1. Es por ello que se corrió dicho modelo ARIMA. En la Tabla 2, se presenta la información respectiva, incluyendo el valor estimado del coeficiente, su error estándar, el estadístico

Tabla 1: Modelos ARIMA y su criterio de información de Akaike

Modelo	AIC	Modelo	AIC	Modelo	AIC	Modelo	AIC
ARIMA (2,2,2)	Inf	ARIMA (15,2,0)	3975490	ARIMA (32,2,0)	3959957	ARIMA (49,2,0)	3954518
ARIMA (0,2,0)	4313107	ARIMA (16,2,0)	3973933	ARIMA (33,2,0)	3959394	ARIMA (50,2,0)	3954428
ARIMA (1,2,0)	4149768	ARIMA (17,2,0)	3972490	ARIMA (34,2,0)	3958968	ARIMA (51,2,0)	3954289
ARIMA (0,2,1)	Inf	ARIMA (18,2,0)	3971381	ARIMA (35,2,0)	3958260	ARIMA (52,2,0)	3954199
ARIMA (2,2,0)	4090851	ARIMA (19,2,0)	3969677	ARIMA (36,2,0)	3957947	ARIMA (53,2,0)	3953946
ARIMA (3,2,0)	4057536	ARIMA (20,2,0)	3968703	ARIMA (37,2,0)	3957578	ARIMA (54,2,0)	3953798
ARIMA (4,2,0)	4039292	ARIMA (21,2,0)	3967392	ARIMA (38,2,0)	3957286	ARIMA (55,2,0)	3953594
ARIMA (5,2,0)	4023875	ARIMA (22,2,0)	3966348	ARIMA (39,2,0)	3956736	ARIMA (56,2,0)	3953482
ARIMA (6,2,0)	4014710	ARIMA (23,2,0)	3965412	ARIMA (40,2,0)	3956518	ARIMA (57,2,0)	3953304
ARIMA (7,2,0)	4005966	ARIMA (24,2,0)	3964831	ARIMA (41,2,0)	3956237	ARIMA (58,2,0)	3953191
ARIMA (8,2,0)	4000127	ARIMA (25,2,0)	3964017	ARIMA (42,2,0)	3955952	ARIMA (59,2,0)	3952882
ARIMA (9,2,0)	3994560	ARIMA (26,2,0)	3963406	ARIMA (43,2,0)	3955778	ARIMA (60,2,0)	3952789
ARIMA (10,2,0)	3990225	ARIMA (27,2,0)	3962787	ARIMA (44,2,0)	3955613	**ARIMA (61,2,0)**	**3952739**
ARIMA (11,2,0)	3985969	ARIMA (28,2,0)	3962243	ARIMA (45,2,0)	3955392	ARIMA (62,2,0)	Inf
ARIMA (12,2,0)	3983383	ARIMA (29,2,0)	3961497	ARIMA (46,2,0)	3955266	ARIMA (61,2,1)	Inf
ARIMA (13,2,0)	3980524	ARIMA (30,2,0)	3960995	ARIMA (47,2,0)	3955014	ARIMA (2,1,0)	4090862
ARIMA (14,2,0)	3978143	ARIMA (31,2,0)	3960289	ARIMA (48,2,0)	3954772	ARIMA (62,2,1)	Inf

Tabla 2: Coeficientes (Coef) ARIMA.

Coef	Valor Estimado	Error Estándar	Estadístico Z	P-value	Coef	Valor Estimado	Error Estándar	Estadístico Z	P-value
ar1	-1.04164	0.0014325	-727.1488	< 2.2e-16	ar31	-0.5577189	0.0061985	-89.9764	<2.2e-16
ar2	-1.01699	0.0020679	-491.8074	< 2.2e-16	ar32	-0.5338163	0.0061951	-86.1672	< 2.2e-16
ar3	-1.00674	0.0025282	-398.2036	< 2.2e-16	ar33	-0.5199244	0.006183	-84.089	< 2.2e-16
ar4	-0.99238	0.0029088	-341.1652	< 2.2e-16	ar34	-0.4980559	0.0061651	-80.7858	< 2.2e-16
ar5	-0.99173	0.0032351	-306.5525	< 2.2e-16	ar35	-0.4784613	0.0061401	-77.9243	< 2.2e-16
ar6	-0.97423	0.0035301	-275.9748	< 2.2e-16	ar36	-0.45035	0.0061084	-73.7269	< 2.2e-16
ar7	-0.97054	0.0037919	-255.9492	< 2.2e-16	ar37	-0.4337656	0.0060679	-71.4849	< 2.2e-16
ar8	-0.9531	0.0040338	-236.281	< 2.2e-16	ar38	-0.4149177	0.0060221	-68.8994	< 2.2e-16
ar9	-0.94385	0.0042523	-221.9621	< 2.2e-16	ar39	-0.3974993	0.0059688	-66.596	< 2.2e-16
ar10	-0.92652	0.0044551	-207.9677	< 2.2e-16	ar40	-0.3710843	0.0059078	-62.813	< 2.2e-16
ar11	-0.91024	0.0046407	-196.1407	< 2.2e-16	ar41	-0.3558524	0.0058368	-60.9671	< 2.2e-16
ar12	-0.88773	0.0048118	-184.4887	< 2.2e-16	ar42	-0.3374727	0.0057594	-58.5953	< 2.2e-16
ar13	-0.87756	0.0049670	-176.6800	< 2.2e-16	ar43	-0.3182995	0.005672	-56.1176	<2.2e-16
ar14	-0.85902	0.0051119	-168.0428	< 2.2e-16	ar44	-0.3044335	0.0055778	-54.5799	< 2.2e-16
ar15	-0.8403	0.0052446	-160.2202	< 2.2e-16	ar45	-0.2901119	0.0054762	-52.9773	< 2.2e-16
ar16	-0.81408	0.0053667	-151.6918	< 2.2e-16	ar46	-0.2730868	0.0053667	-50.8854	< 2.2e-16
ar17	-0.80020	0.0054761	-146.1251	< 2.2e-16	ar47	-0.260087	0.0052447	-49.5909	< 2.2e-16
ar18	-0.78646	0.0055777	-140.9999	< 2.2e-16	ar48	-0.2403391	0.0051119	-47.0152	< 2.2e-16
ar19	-0.77528	0.0056720	-136.6848	< 2.2e-16	ar49	-0.2201391	0.004967	-44.3202	< 2.2e-16
ar20	-0.75132	0.0057594	-130.4502	< 2.2e-16	ar50	-0.1997541	0.0048119	-41.5124	< 2.2e-16
ar21	-0.73768	0.0058369	-126.3828	< 2.2e-16	ar51	-0.1876353	0.0046408	-40.4315	< 2.2e-16
ar22	-0.71522	0.0059079	-121.0617	< 2.2e-16	ar52	-0.1725329	0.0044552	-38.7263	< 2.2e-16
ar23	-0.69536	0.0059689	-116.4974	< 2.2e-16	ar53	-0.1595891	0.0042523	-37.5296	< 2.2e-16
ar24	-0.67716	0.0060222	-112.4449	< 2.2e-16	ar54	-0.1378201	0.0040338	-34.1664	< 2.2e-16
ar25	-0.66548	0.006068	-109.6701	< 2.2e-16	ar55	-0.1206184	0.0037919	-31.8097	< 2.2e-16
ar26	-0.64678	0.0061084	-105.8822	< 2.2e-16	ar56	-0.100681	0.0035301	-28.5207	< 2.2e-16
ar27	-0.63116	0.0061402	-102.7923	< 2.2e-16	ar57	-0.0854091	0.003235	-26.4016	< 2.2e-16
ar28	-0.6148	0.0061652	-99.7201	< 2.2e-16	ar58	-0.0665145	0.0029087	-22.8674	< 2.2e-16
ar29	-0.5983	0.0061831	-96.7634	< 2.2e-16	ar59	-0.0505129	0.002528	-19.9811	< 2.2e-16
ar30	-0.57564	0.0061952	-92.918	< 2.2e-16	ar60	-0.0248847	0.0020676	-12.0354	< 2.2e-16
					ar61	-0.0104799	0.0014321	-7.3178	< 2.52e-13

Z y el p-value. Con el p-value se puede saber si los coeficientes son estadísticamente significativos, individualmente. Como es bien sabido, el p-value tiene que ser menor que el valor de nuestro error tipo I, que en este caso se asumió del 5 %; analizando la tabla, el p-value es mucho menor que nuestro error tipo I, por lo que todos los coeficientes son estadísticamente significativos individualmente a un nivel del 0.1 %, que es aún más bajo que lo requerido.

Para que el modelo ARIMA sea el mejor posible, este tiene que cumplir con dos requisitos fundamentales: (i) que la media de los residuos sea cero y (ii) que los residuos no estén autocorrelacionados. En la Figura 9 se pueden apreciar tres gráficas: la superior, muestra cómo se desarrollan los residuos; aquí se verifica que estos tienen una media cero. La segunda, que está en la esquina inferior izquierda, muestra la ACF, la cual en los primeros 15 lags muestra que los residuos no están autocorrelacionados; no obstante, de ahí en adelante, se muestra una clara autocorrelación en sentido negativo y que cada vez va en aumento, lo cual es un resultado nada intuitivo y de cierta manera arroja múltiples preguntas que podrán ser atendidas en una subsecuente investigación. La tercera gráfica muestra cómo están distribuidos los residuos, donde la línea naranja traza la forma de la distribución de estos, esperando, en un caso utópico, que esta fuera una distribución normal; sin embargo, como se está trabajando con un activo extremadamente volátil, esto no es así. De hecho, se asemeja más a una delta de Dirac para este caso en particular.

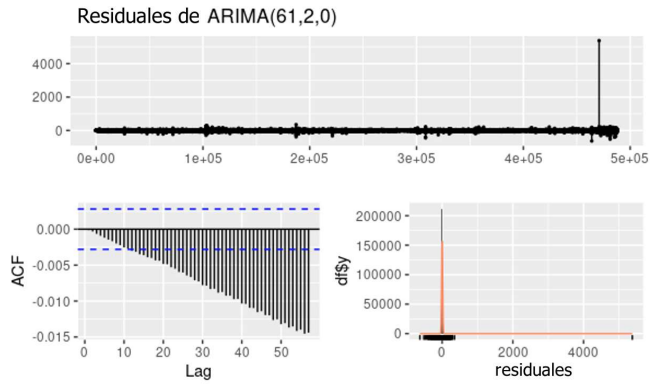


Figura 9: Análisis de los residuales ARIMA (61,2,0).

5. Análisis posterior

Una vez con el modelo ARIMA (61,2,0), se realizó un pronóstico de los siguientes veinte (y últimos) movimientos de 2020. Los resultados fueron bastante cercanos a los valores reales. Dentro del pronóstico, sólo dos valores se salieron de las bandas de error teórico (a un 95 % de confianza), como se puede apreciar en la Figura 10. No obstante, obteniendo un error con los valores reales, se obtuvo que ninguna predicción difirió en más del 1 %; de hecho, el valor más errado varió en 0.48 %. En la Tabla 3, se comparan los valores reales, los pronósticos y su error; este fue obtenido mediante la ecuación 10, dónde P es el pronóstico, V el valor real y E el valor absoluto del error en términos porcentuales.

Tabla 3: Pronóstico vs Valor Real.

Movimiento	Valor Real	Pronóstico	Error
1	29042.53	29046.86	0.01 %
2	29056.19	29044.12	0.04 %
3	29027.68	29042.37	0.05 %
4	28997.53	29040.42	0.15 %
5	28904.58	29038.91	0.46 %
6	28898.45	29036.55	0.48 %
7	28964.38	29034.54	0.24 %
8	28961.01	29032.75	0.25 %
9	28988.99	29031.11	0.15 %
10	29023.61	29028.85	0.02 %
11	29014.32	29027.28	0.04 %
12	29006.61	29025.58	0.07 %
13	29021.56	29023.53	0.01 %
14	29049.51	29021.14	0.10 %
15	29036.10	29019.17	0.06 %
16	29052.02	29016.89	0.12 %
17	29039.53	29014.41	0.09 %
18	29044.79	29010.95	0.12 %
19	29000.12	29008.68	0.03 %
20	28992.79	29005.87	0.05 %

$$E = \left| \frac{P}{V} - 1 \right| \cdot 100. \tag{10}$$

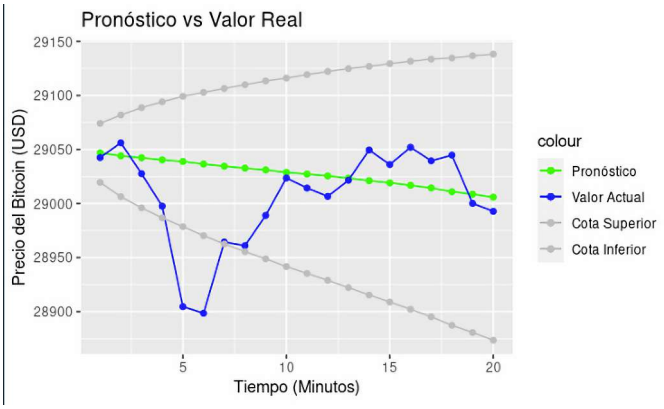


Figura 10: Pronóstico vs Real.

En la Figura 10 se aprecia de una manera visual cómo se comportó el modelo y su pronóstico, siendo la línea verde la predicción, la línea azul los valores reales y las líneas grises las bandas de error teórico al 5 %. En este caso, sólo dos de los veinte pronósticos generados se salieron del margen de error teórico, no obstante, esta infortuna predictiva representa solamente un 0.46 % y 0.48 % de imprecisión con respecto al valor real.

Como se pudo observar, el modelo tuvo una alta precisión a la hora de compararse con los datos reales, a pesar de cómo se comportar sus residuales. El hecho que tenga residuales autocorrelacionados es señal de que hay ciertos términos que debieron incluirse en el modelo final, empero, la capacidad técnica disponible al momento de realizar esta investigación no permitía una corrida tan ambiciosa debido al gran número de términos autorregresivos, así como a la gran cantidad de datos de la muestra;

aunque los resultados obtenidos -sobre todo con respecto a los residuales- arrojan luz para futuras investigaciones.

6. Conclusiones

Esta investigación tiene el propósito de analizar de forma sistemática y organizada las criptomonedas, con especial énfasis en el Bitcoin. En una primera aproximación, se optó por utilizar los bien conocidos modelos ARIMA, y en el desarrollo del mismo se puede extraer lo siguiente:

- El Bitcoin es un activo extremadamente volátil, por lo que modelarlo a través de ARIMA resulta bastante complejo y requiere de un modelo con demasiados términos autorregresivos, sobre todo cuando la muestra es lo suficientemente grande.
- Los residuales del modelo ARIMA muestran autocorrelación, por lo que en la búsqueda de un mejor modelo es necesario mirar hacia otras alternativas como lo son los modelos GARCH (modelo autorregresivo generalizado con heterocedasticidad condicional -por sus siglas en inglés Generalized Autoregressive Conditional Heteroskedasticity-), o métodos mucho más recientes como lo son las redes neuronales y el Machine Learning.
- Si bien el modelo ARIMA tuvo ciertos inconvenientes con sus residuos, mostró una predicción bastante certera comparado con los valores reales; teniendo en su peor predicción un error del 0.48 %.

Agradecimientos

Los autores agradecen al Área Académica de Matemáticas y Física por las facilidades para el uso de instalaciones y en particular del servidor Tollan de la misma.

Referencias

Alahmari, S. A. (2019). Using machine learning arima to predict the price of cryptocurrencies. *ISC Int. J. Inf. Secur.*, pp. 139–144.

Bakar, N. A. y Rosbi, s. (2017). Autoregressive integrated moving average (arima) model for forecasting cryptocurrency exchange rate in high volatility environment: A new insight of bitcointransaction. *International Journal of Advanced Engineering Research and Science*, pp. 130–137.

Bitstamp (2020). Cryptodatadownload. Consultado el 15 de diciembre de 2022, en https://www.cryptodatadownload.com/data/bitstamp/.

INEGI (2022). Empleo y ocupación: Población económicamente activa. Consultado el 10 de diciembre de 2022, en https://www.inegi.org.mx/temas/empleo/.

Levendis, J. (2019). *Time Series Econometrics*. Springer, New York.

Roy, S., Nanjiba, S., y Chakrabarty, A. (2018). Bitcoin price forecasting using time series analysis. *2018 21st International Conference of Computer and Information Technology (ICCIT)*, pp. 1–5.